Convergence Analysis of a Multiobjective Artificial Immune System Algorithm

Mario Villalobos-Arias[1,3], Carlos A. Coello Coello[2] and Onésimo Hernández-Lerma[1]

[1]CINVESTAV-IPN
Department of Mathematics
A. Postal 14-740
México, D.F. 07000, MEXICO
[2]CINVESTAV-IPN
Evolutionary Computation Group
Depto. de Ingeniería Eléctrica
Sección de Computación
Av. Instituto Politécnico Nacional No. 2508
Col. San Pedro Zacatenco
México, D. F. 07300, MEXICO
[3] Permanent Address:
Escuela de Matemática
Universidad de Costa Rica
San José, Costa Rica
{mava,ohernand}@math.cinvestav.mx
ccoello@cs.cinvestav.mx
mvillalo@cariari.ucr.ac.cr

Abstract. This paper presents a mathematical proof of convergence of a multi-objective artificial immune system algorithm (based on clonal selection theory). An specific algorithm (previously reported in the specialized literature) is adopted as a basis for the mathematical model presented herein. The proof is based on the use of Markov chains.

1 Introduction

Despite the considerable amount of research related to artificial immune systems in the last few years [3, 10], there is still little work related to issues as important as mathematical modelling (see for example [13, 11]). Other aspects, such as convergence, have been practically disregarded in the current specialized literature.

Problems with several (maybe conflicting) objectives tend to arise naturally in most domains. These problems are called "multiobjective" or "vector" optimization problems, and have been studied in Operations Research where a number of solution techniques have been proposed [9]. It was until relatively recently that researchers became aware of the potential of population-based heuristics such as artificial immune systems in this area [6, 1]. The main motivation for using population-based heuristics (such as artificial immune systems) in solving multiobjective optimization problems is because such a population makes possible to deal simultaneously with a set of possible solutions (the so-called population) which allows us to find several members of the Pareto optimal set in a single run of the algorithm, instead of having to perform a series of separate runs as in the case of the traditional mathematical programming techniques [9]. Additionally, population-based heuristics are less susceptible to the shape or continuity of the Pareto front (e.g., they can easily deal with discontinuous and concave Pareto fronts), whereas these two issues are a real concern for mathematical programming techniques [5, 1].

This paper deals with convergence analysis of an artificial immune system algorithm used for multiobjective optimization.

The remainder of this paper is organized as follows. Section 2 briefly describes the general multiobjective optimization problem and introduces the basic definitions adopted in this paper. In Section 3, we briefly describe the specific algorithm adopted for developing our mathematical model of convergence. Then, in Section 4, we present our main results. A mathematical proof of such results is presented in Section 5. Finally, our conclusions and some possible paths for future research are presented in Section 6.

2 The multiobjective optimization problem

To compare vectors in $\mathbb{R}^d$ we will use the standard *Pareto order* defined as follows.

If $\boldsymbol{u} = (u_1, u_2, \ldots, u_d)$ and $\boldsymbol{v} = (v_1, v_2, \ldots, v_d)$ are vectors in $\mathbb{R}^d$, then

$$\boldsymbol{u} \preceq \boldsymbol{v} \iff u_i \leq v_i \;\forall\, i \in \{1, \ldots, d\}.$$

This relation is a *partial order*. We also write $\boldsymbol{u} \prec \boldsymbol{v} \iff \boldsymbol{u} \preceq \boldsymbol{v}$ and $\boldsymbol{u} \neq \boldsymbol{v}$.

Let X be a set and $F : X \longrightarrow \mathbb{R}^d$ a given vector function with components $f_i : X \longrightarrow \mathbb{R}$ for each $i \in \{1, \ldots, d\}$.

The multiobjective optimization problem (MOP) we are concerned with is to find $x^* \in X$ such that

$$F(x^*) = \min_{x \in X} F(x) = \min_{x \in X} [f_1(x), \ldots, f_n(x)], \tag{1}$$

where the minimum is understood in the sense of the Pareto order.

Definition 1:

A point $x^* \in X$ is called a *Pareto optimal solution* for the MOP (1) if there is no $x \in X$ such that $F(x) \prec F(x^*)$.

The set

$$\mathcal{P}^* = \{x \in X \;:\; x \text{ is a Pareto optimal solution}\}$$

is called the *Pareto optimal set*, and its image under F, i.e.

$$F(\mathcal{P}^*) \;:=\; \{F(x) \;:\; x \in \mathcal{P}^*\},$$

is the *Pareto front*.

As we are concerned with the artificial immune system algorithm in which the elements are represented by a string of length l with 0 or 1 in each entry. In the remainder of the paper we will replace X with the *finite* set $\mathbb{B}^l$, where $\mathbb{B} = \{0, 1\}$.

3 The Artificial Immune System Algorithm

For our mathematical model, we will consider the artificial immune system (based on clonal selection theory [4]) for multiobjective optimization proposed in [2]. From here on, we will refer to this approach using the same name adopted by the authors of this algorithm: “Multi-objective Immune System Algorithm” (MISA for short). Next, we will focus our discussion only on the aspects that are most relevant for its mathematical modelling. For a detailed discussion on this algorithm, readers should refer to [2].

MISA is a technique in which there is a population that evolves as follows. The population is divided in two parts, a primary set and a secondary set; the primary set contains the “best” individuals (or elements) of the population. The transition of one population to another is made by means of a mutation rule and a reordering operation. First, several copies of the elements of the primary set are made; then, a “small” mutation (using a parameter or probability p_m) is applied to these copies, while a mutation with the parameter ρ_m is applied to the secondary set. These parameters are positive and less than $1/2$, i.e.

$$p_m, \rho_m \in (0, 1/2)\,. \tag{2}$$

We model this algorithm with a Markov chain $\{X_k : k \geq 0\}$, with state space $S = \mathbb{B}^{nl}$, where $\mathbb{B} = \{0, 1\}$. Hence S is the set of all possible vectors of n individuals each one represented by a string of length l with 0 or 1 in each entry.

In our model, we omitted the clonation stage, and mutation is applied directly to the elements of the primary set. Note that if we use clones the probability of passing from a state to another is increased. Thus, our omission is not relevant to the model, since it does not affect our proof.

The chain’s transition probability is given by

$$P_{ij} = \mathbb{P}(X_{k+1} = j \mid X_k = i).$$

We also write

$$P(i, A) = \mathbb{P}(X_{k+1} \in A \mid X_k = i).$$

Thus the transition matrix is of the form

$$P = (P_{ij}) = RM, \tag{3}$$

where R and M are the transition matrices of reordering and mutation, respectively.

Note that these matrices are stochastic, i.e. $R_{ij} \geq 0$, $M_{ij} \geq 0$ for all i, j, and for each $i \in S$

$$\sum_{s \in S} R_{is} = 1 \;\text{ and }\; \sum_{s \in S} M_{is} = 1. \tag{4}$$

Suppose that the primary set has n_1 individuals, so that the secondary set has $n - n_1$ individuals. Let $i \in S$ be a state (population). Then we can express i as

$$i = (i_1, i_2, \ldots, i_{n_1}, i_{n_1+1}, \ldots, i_n),$$

where each i_s is a string of length l of 0’s and 1’s.

3.1 The Mutation Probability

In order to calculate the mutation probability from the state i to state j we use that the individual i_s is transformed into the individual j_s applying uniform mutation (i.e. each entrance of i_s is transformed into the corresponding one of j_s with probability p_m or ρ_m), as in the following scheme.

$$
\begin{array}{c|c|c|c||c|c|c|}
 & 1 & \cdots & n_1 & n_1+1 & \cdots & n \\
\cline{2-7}
i & i_1 & \cdots & i_{n_1} & i_{n_1+1} & \cdots & i_n \\
\cline{2-7}
\multicolumn{1}{c}{\text{mutation}} & \multicolumn{1}{c}{\downarrow} & \multicolumn{1}{c}{\cdots} & \multicolumn{1}{c}{\downarrow} & \multicolumn{1}{c}{\downarrow} & \multicolumn{1}{c}{\cdots} & \multicolumn{1}{c}{\downarrow} \\
\cline{2-7}
j & j_1 & \cdots & j_{n_1} & j_{n_1+1} & \cdots & j_n \\
\cline{2-7}
\end{array}
$$

Thus, for each individual in the primary set of the population, the mutation probability can be computed as

$$p_m^{H(i_s,j_s)}(1-p_m)^{l-H(i_s,j_s)} \ \ \forall s \in \{1,\ldots,n_1\},$$

where $H(i_s, j_s)$ is the Hamming distance between i_s and j_s. For the secondary set we have

$$\rho_m^{H(i_s,j_s)}(1-\rho_m)^{l-H(i_s,j_s)} \ \ \forall s \in \{n_1+1,\ldots,n\}.$$

Hence the mutation probability from i to j is:

$$M_{ij} = \prod_{s=1}^{n_1} p_m^{H(i_s,j_s)}(1-p_m)^{l-H(i_s,j_s)} \prod_{s=n_1+1}^{n} \rho_m^{H(i_s,j_s)}(1-\rho_m)^{l-H(i_s,j_s)} \tag{5}$$

3.2 Use of Elitism

We say that we are using *elitism* in an algorithm, in particular in MISA, if we use an extra set, called the *elite* set (or secondary population, as in [2]), in which we put the "best" elements (nondominated elements of the state in our case) of the primary set. This elite set usually does not participate in the evolution, since it is used only to store the nondominated elements found along the process.

After each transition we have to apply an *elitism operation* that accepts a new state if there is an element in the primary or secondary set that improves some element in the elite set.

If we are using elitism, the representation of the states changes to the following form:

$$\hat{i} = (i^e; i) = (i_1^e, \cdots, i_r^e; i_1, \cdots, i_{n_1}, i_{n_1+1}, \cdots, i_n)$$

where $i_1^e, \cdots, i_r^e$ are the members of the elite set of the state, r is the number of elements in the elite set and we assume that the cardinality of $\mathcal{P}^*$ is greater than r. In addition we assume that $r \leq n$.

Note that in general $i_1^e, \cdots, i_r^e$ are not necessarily the "best" elements of the state $\hat{i}$, but after applying the elitism operation on i^e, they are the "best" elements of the states.

Let $\hat{P}$ be the transition matrix associated with the new states.

Note that if all the elements in the elite set of a state are Pareto optimal then any state that contains an element in the elite set that is not a Pareto optimal will not be accepted, i.e.

$$\text{if } \{i_1^e, \cdots, i_r^e\} \subset \mathcal{P}^* \text{ and } \{j_1^e, \cdots, j_r^e\} \not\subset \mathcal{P}^* \text{ then } \hat{P}_{ij} = 0 \tag{6}$$

4 Main results

Before stating our main results, we introduce the definition of convergence of an algorithm, where we use the following notation. If $V = (v_1, v_2, \ldots, v_n)$ is a vector, we denote by $\{V\}$ the set of entries of V, i.e.

$$\{V\} = \{v_1, v_2, \ldots, v_n\}.$$

Definition 2:

Let $\{X_k : k \geq 0\}$ be the Markov chain associated to an algorithm. We say that the algorithm converges if

$$\mathbb{P}(\{X_k\} \subset \mathcal{P}^*) \to 1 \text{ as } k \to \infty$$

In the case of using elitism, we replace X_k by X_k^e, wich is the *elite set* of the state (i.e. if $X_k = i$ then $X_k^e = i^e$)

Theorem 1:

Let P be the transition matrix of MISA. Then, P has a stationary distribution π such that

$$||P^k - \pi|| \leq \left(1 - 2^{nl} p_m^{n_1 l} \rho_m^{(n-n_1)l}\right)^k \quad \forall k = 1, 2, \ldots. \tag{7}$$

Moreover, π has all its entries positive.

In spite of this convergence result, the convergence of MISA to the Pareto optimal set cannot be guaranteed.

In fact, from Theorem 1 and using the fact that π has all entries positive we will immediately deduce the following fact.

Corollary 1:

MISA does not converge.

To ensure convergence of MISA we need to use elitism.

Theorem 2:

The elitist version of MISA *does* converge.

5 Proof of the results

We first recall some standard definitions and results.

Definition 3:

A nonnegative matrix P is said to be *primitive* if there exists a $k > 0$ such that the entries of P^k are all positive.

Definition 4:

A Markov chain $\{X_k : k \geq 0\}$ with transition matrix P, it is said to satisfy a *minorization condition* if there is a pair (β, μ) consisting of a positive real number β and a probability distribution μ on S, and such that

$$P(i, A) \geq \beta\mu(A) \ \ \forall i \in S, \forall A \subseteq S.$$

The following result gives an upper bound on the convergence rate of a Markov chain that satisfies a minorization condition.

Lemma 1:

Cosider a Markov chain $\{X_k : k \geq 0\}$ with transition matrix P and suppose that it satisfies a minorization condition (β, μ). Then P has a unique stationary distribution π. Moreover for any initial distribution we have

$$||P^k - \pi|| \leq (1-\beta)^k \ \ \forall k = 1, 2, \ldots.$$

Proof see for example [7, pp. 56,57]

We will use the next result to show the existence of the stationary distribution in Theorem 1.

Lemma 2:

Let P be a stochastic primitive matrix. Then, as $k \to \infty$, P^k converges to a stochastic matrix $P^\infty = \mathbf{1}' p^\infty$, where $\mathbf{1}'$ is a column vector of 1's and $p^\infty = p^0 \lim_{k\to\infty} P^k = p^0 P^\infty$ has positive entries and it is unique, independently of the initial distribution p^0.
Proof [8, p. 123]

The next lemma will allow us to use either Lemma 1 or Lemma 2.

Lemma 3:

Let P be the transition matrix of MISA. Then

$$\min_{i,j\in S} P_{ij} = p_m^{n_1 l} \rho_m^{(n-n_1)l} > 0 \ \ \forall i, j \in S, \tag{8}$$

and therefore P is primitive. Moreover, P satisfies a minorization condition (β, μ) with

$$\beta = 2^{nl} p_m^{n_1 l} \rho_m^{(n-n_1)l}, \quad \mu(A) = \frac{|A|}{2^{nl}}, \ \ \forall A \subset S \tag{9}$$

where $|A|$ is the cardinality of A.

Proof

By (2) we have

$$p_m < \frac{1}{2} < 1 - p_m, \quad \rho_m < \frac{1}{2} < 1 - \rho_m.$$

Thus, from (5),

$$\begin{aligned} M_{ij} &= \prod_{s=1}^{n_1} p_m^{H(i_s,j_s)}(1-p_m)^{l-H(i_s,j_s)} \prod_{s=n_1+1}^{n} \rho_m^{H(i_s,j_s)}(1-\rho_m)^{l-H(i_s,j_s)} \\ &> \prod_{s=1}^{n_1} p_m^l \prod_{s=n_1+1}^{n} \rho_m^l \\ &= p_m^{n_1 l} \rho_m^{(n-n_1)l} \end{aligned}$$

On the other hand, by (3) and (4)

$$\begin{aligned} P_{ij} &= \sum_{s\in S} R_{is} M_{sj} \\ &\geq p_m^{n_1 l} \rho_m^{(n-n_1)l} \sum_{s\in S} R_{is} \\ &= p_m^{n_1 l} \rho_m^{(n-n_1)l} \; > \; 0, \end{aligned}$$

To verify (8), see that P_{ij} attains the minimum in (8) if i has 0 in all entries and j has 1 in all entries.

Now we will show that the pair (β, μ) given by (9) is a minorization condition for P. Indeed, from (8) we have

$$\begin{aligned} P(i,A) = \sum_{j\in A} P_{ij} &\geq \sum_{j\in A} p_m^{n_1 l} \rho_m^{(n-n_1)l} \\ &= |A| \; p_m^{n_1 l} \rho_m^{(n-n_1)l} \\ &= \frac{|A|}{2^{nl}} \; 2^{nl} \; p_m^{n_1 l} \rho_m^{(n-n_1)l} \\ &= \beta \; \mu(A) \end{aligned}$$

and the desired conclusion follows. ■

Proof of Theorem 1

By Lemma 3, P is primitive. Thus, by Lemma 2, P has a stationary distribution π with all entries positive. Finally, using Lemma 1 and the minorization in (9), we get (7). ■

Before proving Theorem 2 we give some preliminary definitions and results.

Definition 5:

Let X be as in **Definition 1**. We say that X is *complete* if for each $x \in X \setminus \mathcal{P}^*$ there exists $x^* \in \mathcal{P}^*$ such that $F(x^*) \preceq F(x)$.

For instance if X is finite, then X is complete.

Let $i, j \in S$ be two arbitrary states, we say that i *leads* to j, and write $i \to j$, if there exists an integer $k \geq 1$ such that $P_{ij}^k > 0$. If i does not lead to j we write $i \not\to j$.

We call a state i *inessential* if there exists a state j such that $i \to j$ but $j \not\to i$. Otherwise the state i is called *essential*.

We denote the set of essential states by E and the set of inessential states by I. Note that

$$S = E \cup I.$$

We say that P is in *canonical form* if it can be written as

$$P = \begin{pmatrix} P_1 & 0 \\ R & Q \end{pmatrix}.$$

Observe that P can put in this form by reordering the states, that is, the essential states at the beginning and the inessential states at the end. In this case P_1 is the matrix associated with the transitions between essential states, R with transitions between inessential to essential states, and Q with transitions between inessential states. Also note that P^k has a Q^k in the position of Q in P, i.e.

$$P^k = \begin{pmatrix} P_1^k & 0 \\ R_k & Q^k \end{pmatrix},$$

and R_k is a matrix that depends of P_1, Q and R.

Now we present some facts that will be essential in the proof of Theorem 2.

Lemma 4:

Let P be a stochastic matrix, and let Q be the submatrix of P associated with transitions between inessential states. Then, as $k \to \infty$, $Q^k \to 0$ elementwise geometrically fast.

Proof See, for instance, [12, p.120]. ■

As a consequence of Lemma 4 we have the following.

Corollary 2:

If $\{X_k : k \geq 0\}$ is a Markov chain, then

$$\mathbb{P}(X_k \in I) \to 0 \text{ as } k \to \infty,$$

independently of the initial distribution.

Proof

For any initial distribution vector p_0, let $p_0(I)$ be the subvector which corresponds to the inessential states. Then, by Lemma 4,

$$\mathbb{P}(X_k \in I) = p_0(I)' Q^k \mathbf{1} \to 0 \text{ as } k \to \infty. \qquad \blacksquare$$

Proof of Theorem 2

By Corollary 2, it suffices to show that the states that contain elements in the elite set that are not Pareto optimal are inessential states. To this end, first note that $X = \mathbb{B}^l$ is complete because it is finite.

Now suppose that there is a state $\hat{i} = (i^e; i)$ in which the elite set contain elements $i^e_{s_1}, \ldots, i^e_{s_k}$ that are not Pareto optimal. Then, as X is complete, there are elements, say $j^e_{s_1}, \ldots, j^e_{s_k} \in \mathcal{P}^*$, that dominate $i^e_{s_1}, \ldots, i^e_{s_k}$, respectively.

Take $\hat{j} = (j^e; j)$ such that all Pareto optimal points of i^e are in j^e and replace the other elements of i^e with the corresponding $j^e_{s_1}, \ldots, j^e_{s_k}$. Thus all the elements in j^e are Pareto optimal.

Now let

$$j = (j^e_1, \ldots, j^e_r, \underbrace{i^e_{s_1}, \ldots, i^e_{s_1}}_{n-r})$$

By Lemma 3 we have $i \to j$. Hence, with positive probability we can pass from (i^e, i) to (i^e, j), and then we apply the elitist operation to pass from (i^e, j) to (j^e, j). This implies that $\hat{i} \to \hat{j}$. On the other hand, using (6) $\hat{j} \not\to \hat{i}$. Therefore $\hat{i}$ is an essential state.

Finally, from Corollary 2 we have

$$\mathbb{P}(\{X^e_k\} \subset \mathcal{P}^*) = \mathbb{P}(X_k \in E) = 1 - \mathbb{P}(X_k \in I) \to 1 - 0 = 1 \text{ as } k \to \infty.$$

This completes the proof of Theorem 2. $\blacksquare$

6 Conclusions and Future Work

We have presented a proof of convergence for the multiobjective artificial immune system algorithm presented in [2] and called MISA by its authors. The theoretical analysis of the approach indicates that the use of elitism (which is represented in the form of an *elite set* in the case of multiobjective optimization) is necessary to guarantee convergence. To the authors' best knowledge, this is the first time that this sort of mathematical proof of convergence is presented for a multiobjective artificial immune system.

As part of our future work, we plan to extend our theoretical analysis to other types of artificial immune systems [3]. We are also interested in defining a more general framework for proving convergence of heuristics that operate with base on a mutation operator. Such a framework would allow to prove convergence of a family of heuristics that comply with a certain (minimum) set of requirements, rather than having to devise a specific proof for each of them.

Acknowledgments

The first author acknowledges support from the Universidad de Costa Rica through a scholarship to pursue graduate studies at the Department of Mathematics of CINVESTAV-IPN. The second author acknowledges support from NSF-CONACyT project No. 42435-Y. The last author acknowledges partial support from CONACyT grant 37355-E.

References

1. Carlos A. Coello Coello, David A. Van Veldhuizen, and Gary B. Lamont. *Evolutionary Algorithms for Solving Multi-Objective Problems*. Kluwer Academic Publishers, New York, May 2002. ISBN 0-3064-6762-3.
2. Nareli Cruz Cortés and Carlos A. Coello Coello. Multiobjective Optimization Using Ideas from the Clonal Selection Principle. In Erick Cantú-Paz et al., editor, *Genetic and Evolutionary Computation—GECCO 2003. Proceedings, Part I*, pages 158–170. Springer. Lecture Notes in Computer Science Vol. 2723, July 2003.
3. Dipankar Dasgupta, editor. *Artificial Immune Systems and Their Applications*. Springer-Verlag, Berlin, 1999.
4. Leandro Nunes de Castro and F. J. Von Zuben. Learning and Optimization Using the Clonal Selection Principle. *IEEE Transactions on Evolutionary Computation*, 6(3):239–251, 2002.
5. Kalyanmoy Deb. *Multi-Objective Optimization using Evolutionary Algorithms*. John Wiley & Sons, Chichester, UK, 2001. ISBN 0-471-87339-X.
6. Prabhat Hajela and Jun Sun Yoo. Immune Network Modelling in Design Optimization. In D. Corne, M. Dorigo, and F. Glover, editors, *New Ideas in Optimization*, pages 167–183. McGraw-Hill, 1999.
7. O. Hernández-Lerma. *Adaptive Markov Control Processes*. Springer-Verlag, New York, 1989.
8. Marius Iosifescu. *Finite Markov Processes and Their Applications*. John Wiley & Sons, Chichester, UK, 1980.
9. Kaisa M. Miettinen. *Nonlinear Multiobjective Optimization*. Kluwer Academic Publishers, Boston, Massachusetts, 1998.
10. Leandro Nunes de Castro and Jonathan Timmis. *An Introduction to Artificial Immune Systems: A New Computational Intelligence Paradigm*. Springer-Verlag, 2002.
11. Rosa Saab, Raúl Monroy, and Fernando Godínez. Towards a model for an immune system. In Carlos A. Coello Coello, Alvaro de Albornoz, Luis Enrique Sucar, and Osvaldo Cairó Battistutti, editors, *MICAI 2002: Advances in Artificial Intelligence. Second Mexican International Conference on Artificial Intelligence*, pages 401–410, Mérida, Yucatán, México, April 2002. Springer. Lecture Notes in Artificial Intelligence. Volume 2313.
12. E. Seneta. *Non-Negative Matrices and Markov Chains*. Springer-Verlag, New York, second edition, 1981.
13. Alexander Tarakanov and Dipankar Dasgupta. A formal model of an artificial immune system. *BioSystems*, 55:151–158, 2000.